

INTISARI

KOMBINASI WEB USAGE MINING DAN WEB CONTENT MINING UNTUK MENGHASILKAN PROFIL NAVIGASI PENGUNJUNG WEBSITE

Oleh

GUNTUR BUDI HERWANTO

13/356443/PPA/04413

World Wide Web saat ini menjadi sumber utama dalam pencarian informasi. Seorang pengguna mengunjungi website yang menyediakan informasi dan melakukan navigasi sesuai dengan topik informasi yang diinginkan. Semua aktivitas pengunjung website terekam dalam sebuah akses log. Akses log ini dapat menjadi sebuah sumber data penting untuk melihat keseluruhan pola navigasi dari para pengunjung website. Konten yang tersedia pada website juga dapat menjadi sumber penting untuk melihat struktur semantik dari keseluruhan konten website. Kedua sumber data ini dapat dikombinasikan untuk mendapatkan pola navigasi yang menarik dari keseluruhan pengunjung website. Pada penelitian ini digunakan pemodelan topik Latent Dirichlet Allocation (LDA) untuk mendapatkan struktur semantik dari keseluruhan konten website. Hasil dari pemodelan topik tersebut dikombinasikan dengan session yang terbentuk dari akses log. Untuk mengelompokkan pola navigasi yang sama digunakan *clustering* dengan menggunakan algoritme SOM. Pola yang sudah terbagi dalam setiap *cluster* kemudian dicari pola yang relevan dengan menggunakan algoritme *sequential pattern mining* Prefixspan. Hasil dari Prefixspan tersebut merupakan profil navigasi dari pengguna website. Pengujian yang dilakukan terhadap sistem yang dibangun menunjukkan sistem dapat mengeluarkan profil yang sesuai untuk sebagian besar pengguna baru.

Kata-kata kunci : website, profil navigasi, clustering, pemodelan topik, pattern mining

ABSTRACT

COMBINATION OF WEB USAGE MINING AND WEB CONTENT MINING FOR GENERATING WEBSITE USER NAVIGATION PROFILE

By

GUNTUR BUDI HERWANTO

13/356443/PPA/04413

The World Wide Web is now the primary source for information discovery. A user visit websites that provide information and browse the particular information in accordance to their topic interest. All the activities of the visitors recorded in an access log. Access logs can be an important source for finding useful navigational patterns and visitor behaviour. Website content can also be an important source to see the overall semantic structure of the website's content. Both of these data sources can be combined to obtain an interesting navigational pattern of overall visitors. In this research, Latent Dirichlet Allocation (LDA) used to get the semantic structure of the entire content of the website. Results from LDA combined with a session that has been formed from the access log. To group the same navigation pattern, clustering algorithm, Self Organizing Map is used. Navigational patterns that has been grouped within each cluster become a source for sequential pattern mining algorithm, Prefixspan, to get some relevant navigational patterns. Evaluation shows that system can make a navigational profile that match overall navigational profile for new users.

Keywords : website, navigation profile, clustering, topic modelling, pattern mining